Appendix:

A. PROOF

Here we will present the proof of Theorem 1. The proof of Theorem 2 is similar to the proof presented here, but is omitted due to want of space.

In the sequel, ε is used to denote either the empty string or the empty expression. Its intended usage should be clear from the context. The notation α^k, where α is a string and k an integer, is used to represent the string obtained by repeating k times the string α. In particular, $\alpha^0 = \varepsilon$.

Theorem 1 The consistent PAE problem is NP-complete.

Proof. Let *POS* and *NEG* be two sets of strings. Deciding whether or not a string is accepted by a PAE can be done in polynomial time. The size of the shortest PAE that is consistent with respect to $\langle POS, NEG \rangle$ is bounded by the sum of the lengths of the strings in POS. Therefore, this problem is in NP.

To prove that this problem is NP-hard, SAT is reduced to the problem. Assume the alphabet $\Sigma = \{\$, 0, 1\}$.

Let F be a propositional formula in conjunctive normal form with clauses $C_1, C_2, \ldots, C_m$ and variables $V_1, V_2, \ldots, V_n$.

For $1 \leq i \leq m$ and $1 \leq j \leq n$, let us define:

$$F_{ij} \begin{cases} \$10, \text{if } V_j \text{ appears positively in } C_i; \\ \$01, \text{if } V_j \text{ appears negatively in } C_i; \\ \$00, \text{if } V_j \text{ does not appear in } C_i. \end{cases}$$

In a string \$01 and \$10 can be used to represent the logical values true and false, respectively. Thus for all $1 \leq i \leq m$, the string $F_{i1}F_{i2}...F_{in}$ encodes the only assignment of truth values to the variables, $V_1, V_2, ..., V_n$, which makes the clause C_i false. Moreover, define:

$$POS = \{(\$0)^n, (\$1)^n\}$$

$$NEG = N_1 \cup N_2 \cup N_3$$

$$N_1 = \{\$^{n+1}, 0\$^n, 1\$^n\}$$

$$N_2 = \{\$^k010\$^{n-k}, \$^k101\$^{n-k} \mid 1 \leq k \leq n\}$$

$$N_3 = \{F_{i1}F_{i2}...F_{in} \mid 1 \leq i \leq m\}$$

The formula F is satisfiable if there is a PAE that is a consistent with respect to $\langle POS, NEG \rangle$.

Two PAEs, E_t = \$0*1* and E_f = \$1*0*, can be used to represent the logical values true and false, respectively. Given an assignment of truth values to the variables, $V_1, V_2, ..., V_n$, in the formula F, a PAE can be constructed,

$$E_j \begin{cases} E_t, \text{if the truth value assinged to } V_j \text{ is true;} \\ E_f, \text{if the truth value assinged to } V_j \text{ is false.} \end{cases}$$

So if the formula F is satisfiable, then there needs to be an assignment of truth values to the variables, $V_1, V_2, \ldots, V_n$, which satisfies F. It can be shown that if a PAE, E, is constructed as defined above, then E is consistent with respect to $\langle POS, NEG \rangle$.

Now suppose that there is a PAE, E, which is consistent with respect to $\langle POS, NEG \rangle$. Then it follows that $L(E) \supseteq POS$ and $L(E) \cap NEG = \emptyset$. Assuming that E is in a compact form in which the consecutive occurrences of 0* or 1* are collapsed into one, since the resulted expression will still be equivalent to the original one. For instance, $0*1* is equivalent to $0*0*1*. Since $L(E) \supseteq POS$, a * operator must be attached to every occurrence of 0 and 1 in E. Because $L(E) \cap N_1 = \emptyset$, E needs to have the form of $\$\alpha_1\$\alpha_2 \ldots \$\alpha_n$, where each α_i is a sequence of 0* and 1* only. Moreover, both 0* and 1* must appear at least once in each α_i. Because $L(E) \cap N_2 = \emptyset$, it follows that each α_i is either 0*1* or 1*0*. Therefore, an assignment of truth values to the variables, $V_1, V_2, \ldots, V_n$, can be obtained as defined above. Because $L(E) \cap N_3 = \emptyset$, it can be shown that this assignment needs to satisfy the formula F that is in conjunctive normal form.

Thus, $|POS| + |NEG| = O(mn)$. Therefore the problem is NP-hard.